

This week:

During exam period

Monday	Prepare the examination in the lab.
Tuesday	Have examination in the morning and afternoon.
Wednesday	Have examination in the morning. Finish the course report in the afternoon.
Thursday	Join a report for urban computation in the morning. Read the paper about urban computing.
Friday	Join the group meeting in the morning. Submit the documents to YuQuan.
Weekend	Read paper.

1. PGT: Measuring Mobility Relationship using Personal, Global and Temporal Factors

本文致力于轨迹数据的分析与知识挖掘。想要分析轨迹之间的关系。通过轨迹的碰撞得到两个目标之间的关系。并量化这种关系。

传统的量化方法是根据轨迹的碰撞频率，如果两条轨迹的公共点越多，说明这两条轨迹的对象关系越密切。还有一种度量方法是如果两条轨迹的欧式距离越近，那么说明对象关系越密切。

然而上面描述的两种度量方法都是局限性的，而且是比较单一的。如果两个人在同一个楼里上班，尽管他们完全不认识，他们的分数仍然会很高。如果两个人在一年内不同的城市见面了 3 次，尽管他们关系可能很密切，他们的分数仍然会很低。所以本文考虑并添加了几种关系的其他考量条件，这些考量条件和方法是值得我们参考的。

首先，个人因素。这个个人因素就是考虑关于个人的碰撞几率增加进去。就是如果两个人都仅仅在某几天同时去了某些不同地方，或者说他们两人都不太去一个地方，但是只要去就一起出现在那里，那么他们的量化值应该是很高的（关系很好）。这很容易理解。具体的实现方法就是把他们的轨迹随机打乱后再进行碰撞检测，如果碰撞点几乎没有了，那么符合。

全局因素。就是通过所有的训练数据来训练所有地点的全局指标。如果一个地方很少有人去，那么同时去这里的人的关系密切的可能性就大。如果一个地方流动量很大，那么同时去这里的人的关系密切的可能性就相对要小。

时间因素。这里主要考虑连续时间的轨迹相同。如果一段连续时间的轨迹相同，那么肯定比那些间隔碰面的轨迹认识的可能性更大。当然前提是在一段足够长的时间里。

- **Frequency:** $|E_{ij}|$ (meeting frequency).
- **Personal:** $\max_{e_k \in E_{ij}} \{w_{ij}^p(e_k)\} \times |E_{ij}|$ (meeting frequency weighted by the personal factor only).
- **Global:** $\sum_{e_k \in E_{ij}} w_{ij}^g(e_k)$ (meeting frequency weighted by the global factor only).
- **Temporal:** $\sum_{e_k \in E_{ij}} w_{ij}^t(e_k)$ (meeting frequency weighted by the temporal correlation only).
- **Personal + Global:** $\max_{e_k} \{w_{ij}^p(e_k)\} \times \sum_{e_k} w_{ij}^g(e_k)$ (meeting frequency weighted by both personal factor and global factor).
- **Personal + Global + Temporal:** $\max_{e_k} \{w_{ij}^p(e_k)\} \times \sum_{e_k} (w_{ij}^g(e_k) \times w_{ij}^t(e_k))$ (the PGT model which considers all the factors).

这篇文章给轨迹提取人物关系的方法上有很多贡献，如果以后的工作中有类似需求

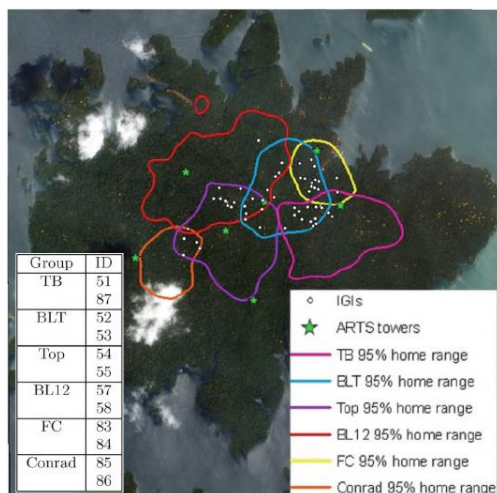
可以借鉴。

2. Attraction and Avoidance Detection from Movements

本文利用了 6 个群落的猩猩移动的数据，进行了基于轨迹的群落关系研究。文章的主要思路是想要挖掘轨迹信息中不同种群之间的吸引和避让关系。主要的思路还是通过碰撞检测的方法。

考虑的因素除了碰撞率之外，碰撞时间和公共活动区域大小也进入了参数的控制范围之内。作者同时对种群之间的这两种行为进行分析，未来想要结合 POI 语义来分析为什么吸引或者避让。是否有其他因素导致间接地吸引和避让。同时对计算的算法进行了时间上的优化。

本文的方法是分析群落之间的关系的一个很好的数据挖掘算法及其实现过程。

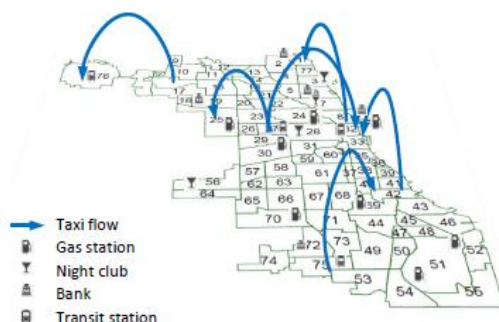


3. Crime Rate Inference with Big Data

这是一篇用于犯罪预测的机器学习文章。主要考量的因素有：传统的考量因素有地理影响因素和该地区内的内部因素。例如人种，犯罪率等等。地理影响因素指的是犯罪率在地理方向上的传播因素。紧挨着的区域会对彼此有影响。

本文新加入的因素有两个，一个是 POI 数据的学习。从样本中抽取出各个地区的 POI 种类数据，然后根据犯罪率赋值 POI。例如酒吧犯罪率低等等。

另一个是交通的影响，根据拿到的出租车等数据，上下客联系多的区域也具有跨地理位置的犯罪传播关系。如下图的犯罪传播关系。



文章对每一种因素进行量化之后，进行了和传统方法的比对，准确率有所提升。

4. Semantic Annotation of Mobility Data using Social Media

本文想要解决的问题是轨迹的注释。结合轨迹数据和微博 POI 等数据对轨迹进行注释。因为轨迹数据的误差，不能够精确地知道对象究竟去了什么地方，发生了什么事情。

一种简单的方法就是就近原则，最靠近她的 POI 设定为目的地点。并根据微博的拆分词云等来注释该轨迹的目的地。

Input: Location history of a mobile user				Annotation	
Record ID	Time	Longitude	Latitude	Record ID	Annotations
r ₁	2013-1-20	40.75051	-73.99349	r ₁	madison, garden, rangers, penguins
r ₂	2013-2-10	40.68312	-73.97597	r ₂	nets, barclays, center, nba
r ₃	2013-2-19	40.75051	-73.99349	r ₃	rangers, madison, square, montreal, canadiens

+

Input: Geo-tagged tweets from the crowd

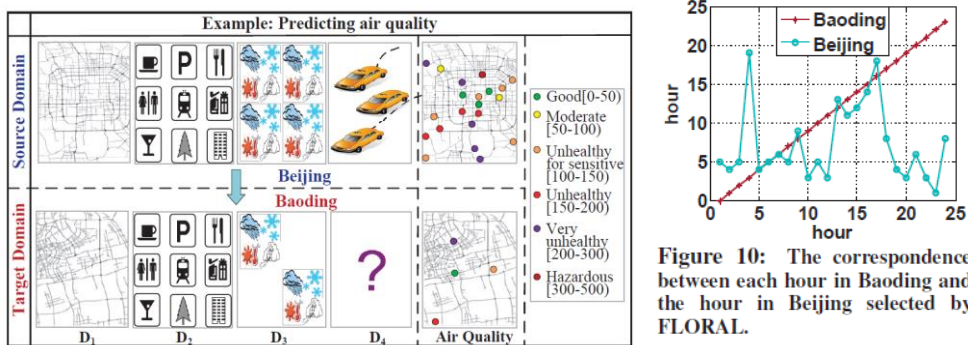
Time	Longitude	Latitude	Tweets
2013-1-20	40.75051	-73.99349	LETS GO RANGERS
2013-1-20	40.61219	-74.15814	I'm at Buffalo Wild Wings
2013-1-20	40.75050	-73.99350	I'm @ Madison Square Garden for Pittsburgh Penguins vs New York Rangers

但是这种方法有一种弊端，如果知道此人一般下午都回去咖啡厅，有一天下午他去了商业区的咖啡馆，但是词云的频率注释则会显示此人下午去了商店逛街等等。本文采用了一种基于高斯函数的 KDE 方法来概率估计人的目的地点，综合此人的运动规律等因素，对人的轨迹数据进行注释分析。

本文的可视化方法比较欠缺，可以思考该方法的数据可视化。

5. Transfer Knowledge between Cities

多源数据在进行数据分析时有一个问题就是数据密度的问题，数据的密度不统一，或者有些数据就是不全面的。本文介绍了一种根据全的数据的城市来转移分析一个数据不全的地区的方法。通过学习语义相关的转移字典来转移具有相关语意标签的数据。



文章首先提取每一种数据源的语义信息并进行量化。然后学习地区的字典矩阵，根据字典矩阵的对比进行从源数据区域到目标数据区域的转换。文章利用北京和包头的数据进行测试，发现如上图，气象方面学习到的数据显示上午和下午时间两地最为接近。所以可以利用数据进行补全分析预测。

这篇文章的启发是根据数据全的地区可以推测一些其他地区的特点。进行数据补全。

Next week:

Poster.

Course report on Tuesday.

Review the paper before Wednesday.

Passport.

Read 5 paper.